

problems developed which had to be corrected. By mid-year, a pair of TIPs had been shipped to Europe, for use in Norway and London. These brought numerous operational problems. For the first time, circuits had to be obtained from a foreign PTT, the circuits were relatively slow at 9.6 Kb, and like Hawaii, these TIPs were on a long spur off the network rather than being doubly connected as IMPs typically are. During 1973, nodes continued to be delivered, but there began to be a low level of switching of node locations, to optimize the use of various IMP configurations and as sites came on and went off the network. Certain improvements were also made to correct problems with the routing algorithm. As 1973 ended the first very distant hosts were connected to the network over telephone lines.

In 1974 there were major efforts to make the network more operationally usable. Subnetwork reliability was improved as was TIP-to-host communication reliability. Methods for providing TIP access control and accounting and partitioning of logical subnetworks of hosts were developed. Methods were developed to selectively reload sections of IMP memory.

In 1975 network development slowed up and the network took on more and more of an operational appearance. Major network developments in 1975 included delivery of the first Pluribus IMP, modification of the IMP and TIP software to support more than

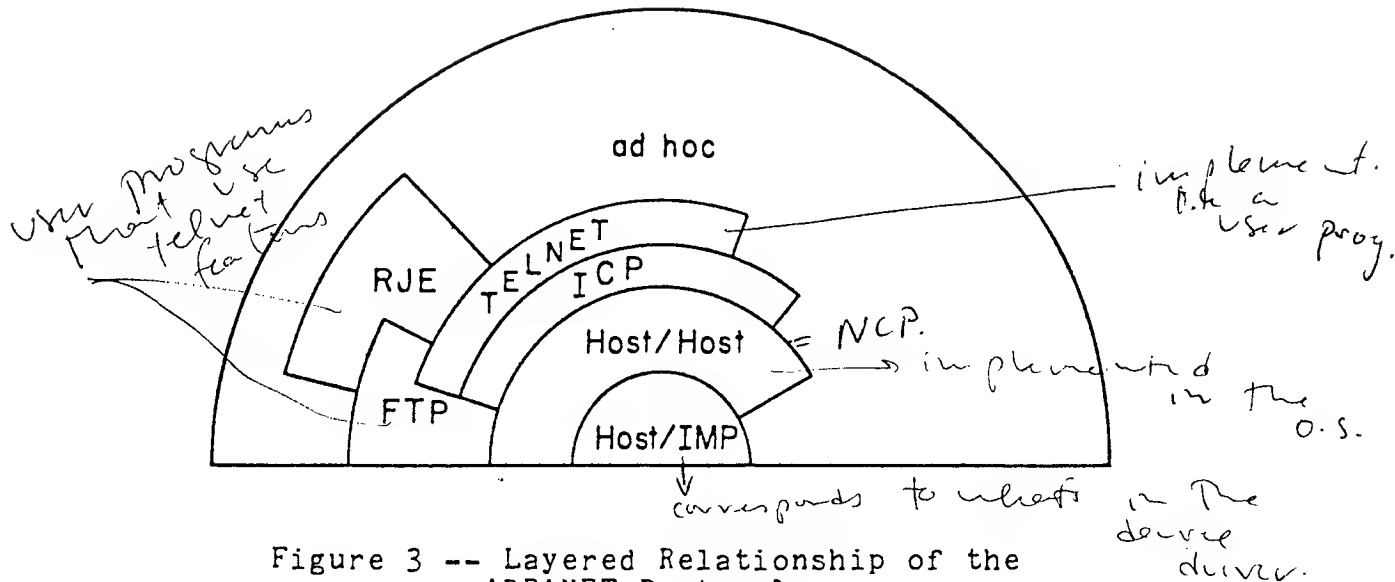
sixty-three IMPs in the network and attachment of the first two Satellite IMPs to the network. By the end of 1975 the network was under DCA management.

Looking back, the subnet development between 1969 and 1975 appears relatively smooth, although there were many times during that period when those intimately involved felt they were trying to solve one crisis or another. The network grew slowly enough, and the basic technology and implementation was flexible and robust enough, that many problems, both major and minor, which naturally cropped up with this new development were for the most part corrected before they obstructed the work of too many users. The fact that the network was also part of an experiment no doubt also made users more tolerant.

1.4.4 Host Protocol Development

Specifications, generally called the IMP-to-host protocol, exist for the physical and logical message transfer between a host and its IMP. This protocol is not sufficient by itself, however, to specify the methods of communication between processes running in two possibly dissimilar hosts. Rather, the processes must have some agreement as to the method of initiating communication, the interpretation of transmitted data, and so forth. Although it would be possible for such agreements to be reached by each pair of hosts (or processes) interested in communication, a more general arrangement is desirable in order to minimize the amount of implementation necessary for network-wide communication. Accordingly, the host organizations formed a group (the Network Working Group or NWG, introduced above) to facilitate an exchange of ideas and to formulate additional specifications for host-to-host communications.

The NWG adopted a "layered" approach to the specification of communications protocols, wherein the higher layers of protocol use the services of lower layers; the advantages and disadvantages of the layered approach are discussed elsewhere in this report. As shown in Figure 3, the lowest layer is the IMP-to-host protocol. The next layer (called the host-to-host layer in the figure) specifies methods of establishing



communications paths between hosts, managing buffer space at each end of a communications path, etc. Next, the Initial Connection Protocol or ICP specifies a standard way for a remote user (or process) to attract the attention of a network host, preparatory to using that host. The ICP provides the analog of the user pressing the attention button at a local terminal on a host. In the next layer is the Telecommunications Network or TELNET protocol, which was designed to support terminal access to remote hosts. TELNET is a specification for a network standard terminal and the protocol for communicating between this standard terminal and a host. The next logical protocol layer consists of function oriented protocols, two of which, File Transfer Protocol (FTP)

and Remote Job Entry protocol (RJE), are shown in the figure. Finally, at any point in the layering process, it is possible to superimpose ad hoc protocols.

In the following subsections we discuss in some detail the events in the evolution of the host-to-host and TELNET protocols, and the events in the evolution of a number of other protocols in somewhat less detail.

USER

USER
PROGRAM or Process

OPEN/close
SEND/RECV

Connection

Operating System

Interface Software
or Device Driver

Messages
with
Headers

Physical
Network

1.4.4.1 Host-to-Host Protocol

The Network Working Group was established in early 1969. By December 1969 an initial host-to-host protocol had been specified which supported communication between a terminal on one host and a process on another host. At a meeting in Salt Lake City in December 1969, the initial protocol specification was described to Lawrence Roberts of DARPA who was unhappy with it because the initial plan would not support transmission of electronic mail over the network. He instructed the Network Working Group to "go back and get it right."

By the spring of 1970 several successive versions of a host-to-host protocol had been developed, and a relatively formal meeting of the NWG was held at UCLA before mid-year at which the latest version of the protocol was described. Reactions to the described protocol were very negative. In June of 1970 there was a series of meetings held at UCLA and Harvard at which people from these two institutions tried finally to settle upon a host-to-host protocol and specify how it should be implemented. In August of 1970 some of the more general (and some thought more exotic) aspects of the host-to-host protocol being considered were ordered dropped from the protocol by Barry Wessler of DARPA, thus administratively clearing away some of those issues which had prevented agreement. The NWG discussion continued at the

1970 Spring Joint Computer Conference; in particular, there was discussion between Crocker and Roberts regarding the formality to be sought for the protocol, and DARPA approvals required, and so forth. Another NWG meeting was held at the Fall Joint Computer Conference in November 1970 in Houston, Texas.

At a NWG meeting held in mid-February 1971 at the University of Illinois, a subcommittee was appointed to look at the host-to-host protocol to see what changes were immediately desirable or necessary. This subcommittee went directly from Illinois to Cambridge, Massachusetts, where it met for two days, wrote an interim report, and then reconvened a month later in Los Angeles. It appears that with the efforts of this committee (known as the "host-to-host protocol glitch cleaning committee") the design of the ARPANET host-to-host protocol was finally coming close to being settled.

At about this same time DARPA was beginning to exert great pressure not only to get the host-to-host protocol settled but also to get it implemented by the hosts. At a NWG meeting at the Spring Joint Computer Conference in Atlantic City in May 1971, Alex McKenzie took on the task of writing a definitive specification of the host-to-host protocol -- not to invent^{or} new protocol, but to write down what had been decided.

In October 1971 the final big NWG meeting was held at M.I.T., and was preceded by a programmers' workshop at which differences in implementations were clarified and eliminated. In January 1972 a McKenzie document describing the protocol was published and the ARPANET host-to-host protocol has remained essentially unchanged since.

TELNET

1.4.4.2 The Evolution of TELNET

Early in the development of the ARPANET it became clear that a major function of the network would be to provide remote use of interactive systems. To allow a user at a terminal (connected to his local host) to control and use a process in a remote host, as if he were a local user of that remote host, a special mechanism was required. The problems to be overcome are legion: for example, the typical host expects its interactive terminals to be physically attached to the individual ports of its hardware terminal scanner rather than logically attached via a multiplexed connection to the network; a given host expects to communicate only with terminals with certain characteristics (e.g., half-duplex, line-at-a-time. physical echo, EBCDIC character set, 134.5 baud) while a remote user's terminal might have completely different characteristics (e.g., full-duplex, character-at-a-time. no character echo, ASCII character set, 300 baud). The TELNET protocol was an attempt to provide the special mechanism necessary to permit such communication.

As early as 1969 a few hosts had been programmed on an ad hoc basis to permit terminal access from another host. In 1971 an NWG subcommittee was formed to consider the general problem of supporting interactive use of arbitrary hosts by users at arbitrary remote terminals. There was great controversy in the

time the revised protocol implementation was scheduled, implementation of the initial version had been completed and host system managers had not budgeted resources for a second implementation; 2) about this time DARPA's research interest in the network was declining and the network was entering a period of status quo operation; 3) despite initial belief that a clean method of phasing over from the initial protocol to the revised protocol existed, none was found by most implementors and consequently most chose to provide a complete implementation of the revised protocol to operate in parallel with the initial protocol; and 4) implementation for the most prevalent user host, the TIP, proved to be very difficult (because of the TIP's limited memory) and time-consuming, thus implicitly relieving pressure on the server hosts to implement the revised protocol. The new TELNET protocol has been the accepted standard for several years, and it is widely implemented and used.

1.4.4.3 The Evolution of the Other Host Protocols

There are several other host protocols the evolution of which should be briefly mentioned.

History Book:

The File Transfer Protocol started out as two protocols, a Data Transfer Protocol and a File Transfer Protocol. To over-simplify, the Data Transfer Protocol was to specify the format of data being transferred and the File Transfer Protocol was to specify how it was transferred. Eventually, the File Transfer Protocol alone was defined with a data portion and a control portion. After the final push to specify the FTP, relatively little additional work was done, consisting only of a little effort to clean up fundamental aspects of the protocol, and a good bit of work reconciling the "reply codes" that different hosts used to indicate FTP-related events.

Before a Remote Job Entry protocol could be defined by the NWG as a whole, UCLA's IBM 360/91 host, a batch oriented host, needed some RJE-like protocol with which to serve a few users who wanted early access to the computing power of that particular host. Thus, led by the UCLA group, a protocol called the Remote Job Service or RJS protocol was defined and implemented. The NWG eventually got around to working on the problem of a Remote Job Entry protocol and undertook a relatively massive effort to define such a protocol. However, by the time the RJE protocol

definition was finished, half a dozen or so hosts had already implemented to interim RJS protocol. Since these included most of the hosts on the network interested in supporting remote batch, there was little incentive for them to implement the new RJE protocol. Thus, today the RJE protocol is carefully specified but to our knowledge is not implemented anywhere, and the RJS protocol prevails.

Moving upward in sophistication, another protocol that was the subject of early discussion was one for graphics. Several versions of a graphics protocol were specified but until recently there was never widespread implementation of any of them. Recently, as part of the ACCAT experiment, an operational graphics protocol has been developed.

In addition to the host-to-host protocol which was finally specified after much iteration, a number of alternative protocols were suggested by various members of the NWG. Before the host-to-host protocol was settled upon, Richard Kaline and David Walden each suggested an alternative protocol. Even after the adoption of the host-to-host protocol, there was some discussion of experiments with a protocol derived from the Walden suggestion. More recently, as part of the DARPA-sponsored National Software works project, Robert Thomas, Stuart Schaffner, and their colleagues have designed and implemented a host-to-host

1.4.3 Subnet Development

The first four IMPs were developed and installed on schedule by the end of 1969. No sooner were these IMPs in the field than it became clear that some provision was needed to connect hosts relatively distant from an IMP (i.e., up to 2000 feet instead of the expected 50 feet). Thus in early 1970 a "distant" IMP/host interface was developed. Augmented simply by heftier line drivers, these distant interfaces made clear that error control was needed on the host/IMP interface. Previously, it had been assumed there would be no errors on such a local hard wired connection.

By mid-year of 1970, a series of network performance tests were being carried out. These uncovered some flaws which were quickly corrected, and some problems which looked more worrisome. Also by mid-year, a rudimentary version of the network control center was established at BBN.

As the year wore on, sites continued to be added to the network, the IMP program continued to be improved, NCC development continued, the first 230.4 Kb circuit was tested between two IMPs, and design for a version of the IMP able to support direct terminal connection was begun. The latter was called the Terminal IMP (TIP).

In 1970 major problems with the IMP flow control and storage allocation techniques were demonstrated. It is interesting that even after these problems were demonstrated (and they were serious enough to completely halt network operation under certain circumstances), the network continued to give adequate service for many, many months while improvements were designed and implemented. The hosts were simply asked to not use the network in the way that caused the subnetwork problems, and the hosts did as they were asked.

(★) About three-quarters of the way through 1971 the first two TIPS were delivered, providing ARPANET access for the first time to users without their own hosts or access to terminals on some other organization's host.

By the beginning of 1972 it was recognized that even the distant version of the IMP/host interface was not sufficient, and design for a IMP/host interface for use over communications circuits was begun. The evolution of the IMP/host interface is worth a little additional comment. The initial bit-serial, asynchronous, non-error-controlled IMP/host interface was essentially specified in the RFQ, in an effort to simplify network connection for the hosts. This non-standard interface may have been of some benefit in simplifying the host connection. However, its greatest virtue was the separation it put between

the IMP and the host. The IMP and host did not have to worry about each other's word size, and they did not have to worry about each other's timing constraints. It seems likely that having to worry about these issues would have delayed network operation. However, this interface also resulted in a hodge-podge of interface variations, each designed for more distant operation than its predecessors, and none except the first was very elegant. For any new network, which need not fear the now proven packet-switching technology, it would clearly be better to use an industry standard communications interface, e.g., HDLC, for every IMP/host connection.

In the first half of 1972 the TIP's capability was expanded to support TIP-to-TIP magnetic tape transfers. While this option was successfully used between two network sites, it was never very elegant. Also, a massive change in the IMP software was undertaken to correct the previously discovered flow control and storage allocation problems. In the second half of the year, the new version of the IMP program was released in many small increments, and the design of a new, ten times more powerful IMP was begun.

★

The beginning of 1973 brought the first satellite link in the network, from California to Hawaii. Also, with network traffic rapidly increasing, a number of subnet reliability

protocol known as MSG. Another protocol, known as TCP, deserves special mention.

Near the time of the formation of the International Network Working Group, as network interconnection began to be of great interest, discussions began on a standard inter-network protocol, particularly one which would correct some of the shortcomings of the ARPANET host-to-host protocol. At the AFIPS 1973 NCC in New York City a meeting was held at which certain ideas for a new host-to-host protocol were discussed. After some additional correspondence, Robert Kahn of DARPA and Vinton Cerf, then of Stanford, got together and designed a protocol known as TCP. Other members of INWG, perhaps not satisfied that TCP represented an international standard, continued developing still another host-to-host protocol (Cerf also participated in this later effort). TCP quickly became DARPA's choice of the host-to-host protocol to be used in situations where the ARPANET host-to-host protocol was insufficient or where inter-networking was required. With DARPA support, several TCP implementations were done and the protocol has come into relatively widespread use within the ARPANET, and its use is still spreading. TCP is scheduled to replace the ARPANET host-to-host protocol throughout the net by 1 January 1983. Meanwhile the host-to-host protocol that the rest of INWG was working on was finished, and documented, just as the PTTs and North American common carriers submitted the X.25

standard to CCITT; so the INWG consensus protocol will most likely play little operational role in the ARPANET or elsewhere.

1.4.5 Network Growth -- A Summary

In the following three subsections we consider three aspects of the growth of the network: the traffic growth, the growth of the network topology, and the increase in the number and type of hosts on the network.

1.4.5.1 Traffic Growth

In early 1973, Roberts presented a curve of average host internode traffic growth for the network* which showed the level of internode network traffic to be increasing at a rate of a factor of ten every ten months. Internode traffic means traffic sent from a host on one node to a host on a different node; i.e., it does not include traffic sent between hosts on the same node. Based on this rapid rate of growth, Roberts predicted the network would run out of capacity in nine months. As shown in the following figure, shortly after Roberts' prediction the rate of internode traffic growth decreased sharply to roughly a factor of two every twenty months. It is interesting to speculate on the reason for this sharp decrease.

* L.G. Roberts, "Network Rationale: A 5-Year Reevaluation," Proceedings COMPCON 1973, February 1973, pp. 3-6.

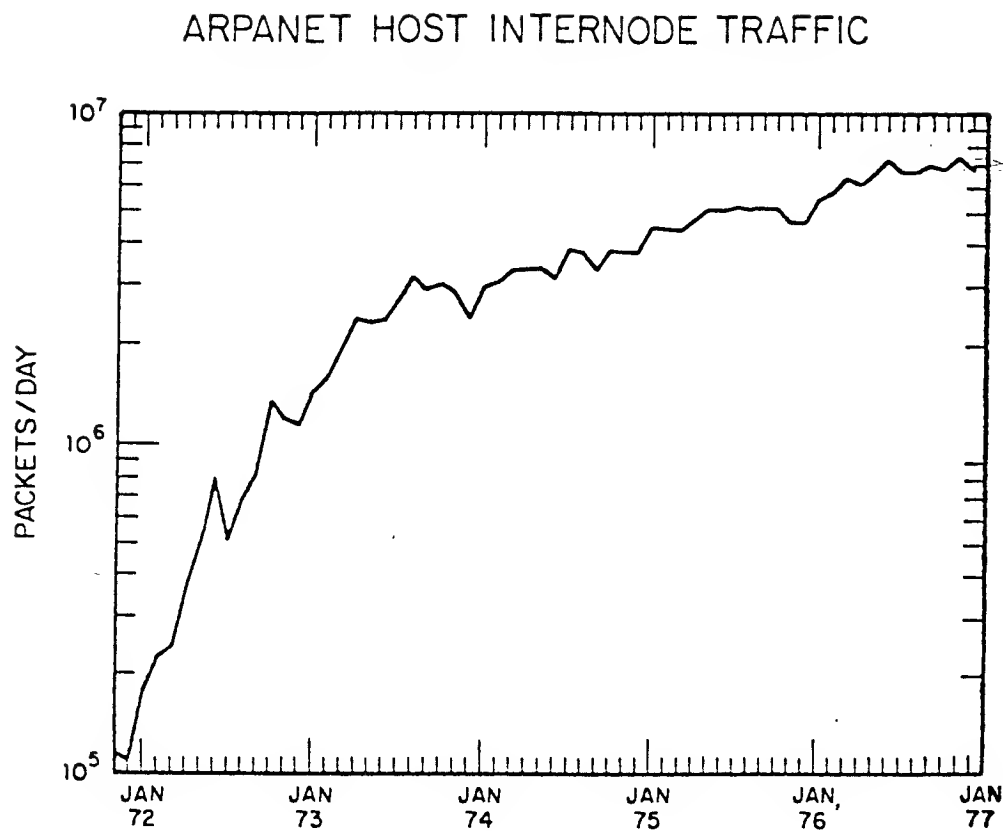


Figure 4 -- Growth in Average Host Internode Traffic

It can be hypothesized that the existing hosts (and the few new hosts that were added) were used remotely more and more and network traffic increased more and more, until the hosts (at least the popular time-sharing hosts) began to run out of capacity; this made it pointless for new remote users to attempt to get service, and resulted, in turn, in a leveling off of network traffic growth. Therefore, instead of the network running out of capacity as predicted by Roberts, it seems that the hosts ran out of capacity while the network still has capacity left.

As already stated, the traffic shown in Roberts' curve and in the figure above includes only internode traffic. There are two reasons for excluding intranode traffic. First, intranode traffic puts a burden on only one node rather than on the network as a whole. Thus when Roberts, for instance, was attempting to calculate the effects of host traffic on network capacity, he naturally excluded intranode traffic. Second, the available intranode traffic statistics include some amount of test traffic being looped from a host through its node and back to the same host, and there is no convenient way to separate this looped test traffic from actual data traffic between two hosts on the same node. It is believed, however, that there is actually a significant amount of real traffic between hosts on the same

node. For instance, Kleinrock reports* that during a week-long measurement, the level of intranode traffic amounted to a daily average of twenty percent of the level of internode traffic, and in some one-hour intervals the intranode traffic level was as much as eighty percent of the internode traffic level. A scan of available long term statistics on inter- and intranode traffic shows that intranode traffic levels have averaged between twenty and forty percent of internode traffic levels. Thus the traffic curve given in the figure above should be scaled up by this factor if all traffic is to be included.

That intranode traffic is a significant portion of all network traffic is interesting and probably indicative of four phenomena. First, the IMP is a handy interhost interface, and once one is installed in a computer center to connect some host onto the network, there is very soon pressure to connect other computers in the computer center to the IMP so that desired communication between the computers is possible. Second, when two computers are connected to the same IMP so they may both communicate with other computers in the network, communication between the two computers themselves comes free and begins to happen even if it was not initially thought to be desired.

* L. Kleinrock and W. Naylor, "On Measured Behavior of the ARPA Network," AFIPS Conference Proceedings, Volume 43, May 1974, pp. 767-780.

Third, the TIP (a host) has been chosen by several sites as the most flexible available terminal multiplexor and TIP-to-host traffic at these sites is likely to be intranode. Fourth, there is a (as yet still weak, but definite) tendency for hosts to be concentrated at a certain site and therefore often on the same IMP. The reason for this tendency is that, while some cynics would have guessed that every computer center manager is trying to build his empire as large as possible, in fact the world of computer center managers appears to include not only managers whose inclinations are as the cynics guessed but also many who dislike running computer centers but do so because they need the service supplied by the computer center. Once the network became available, some sites have arranged with some other sites that one site's computer was moved to a second site, and the second site managed it for the first site which used the computer over the network via a simple terminal concentrator. A further reason for this tendency (for hosts to be clustered) is the economy of scale possible when only one facility and staff is required for the operation of several computers.

1.4.5.2 Topology

The first ARPANET node was installed at the University of California at Los Angeles in late 1969 and the next three nodes were installed in California and Utah.

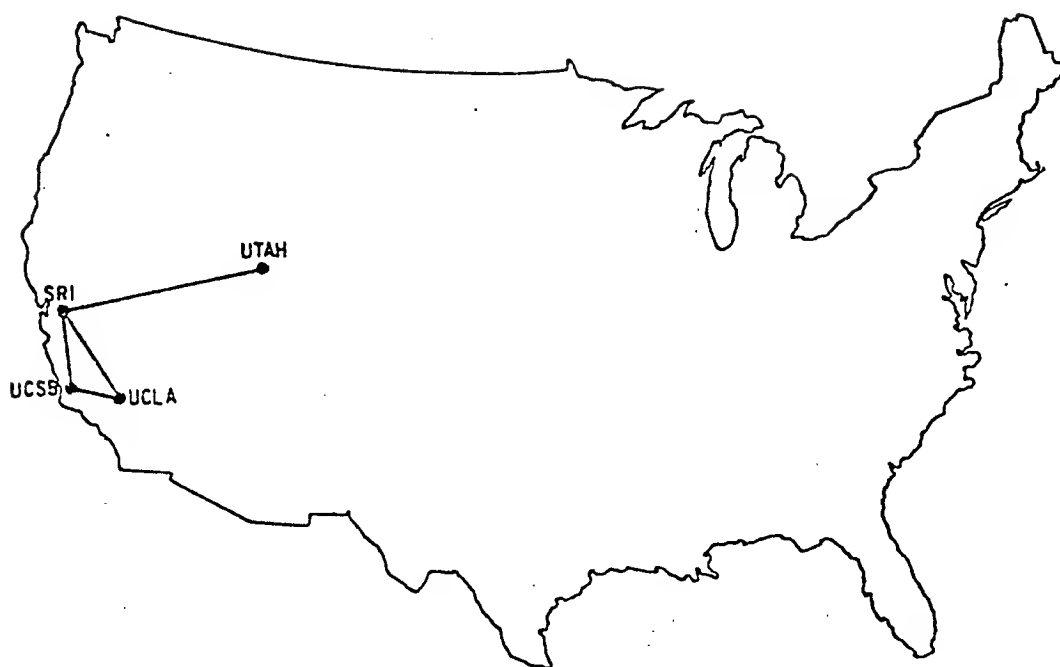


Figure 5: The ARPANET in December 1969

By June 1970 three East Coast and two more West Coast nodes were added, as well as two cross-country lines.

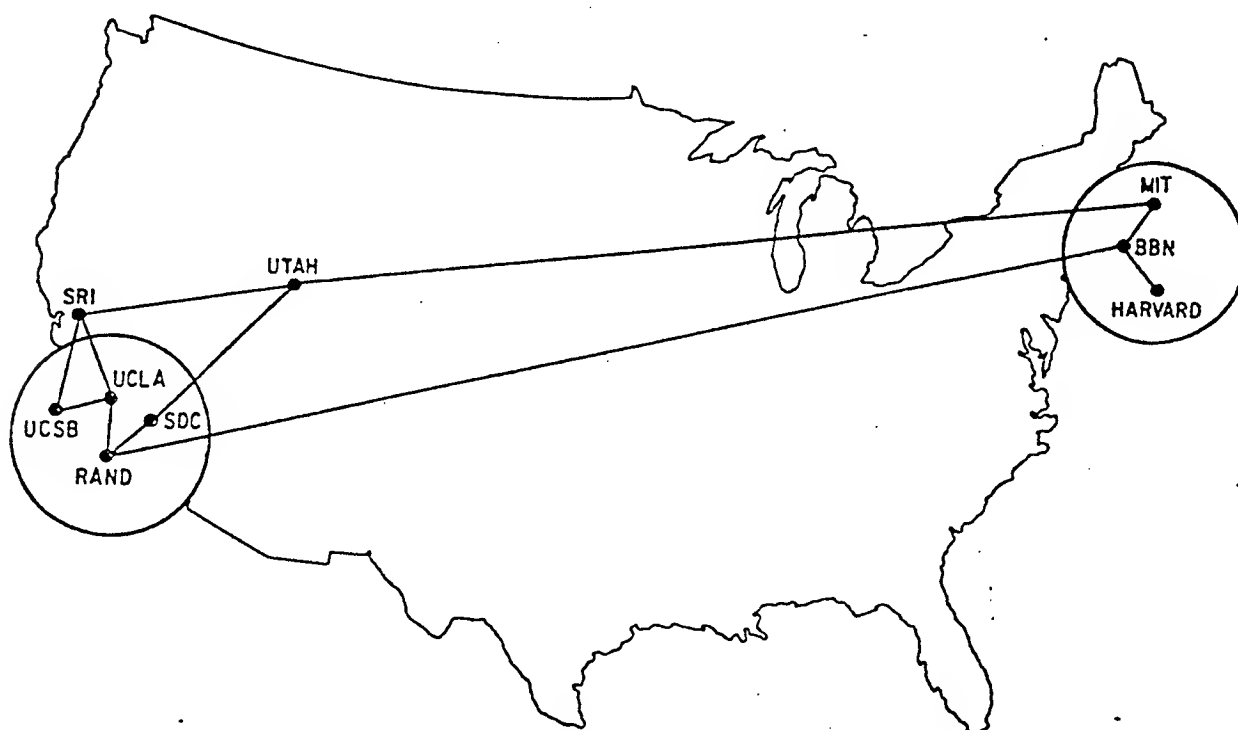


Figure 6: June 1970

IMPs continued to be delivered to the field at an average rate of approximately one per month, so that by late 1970 there were thirteen IMPs installed in the network. The IMPs were all entirely compatible, all being based on the Honeywell 516 computer.

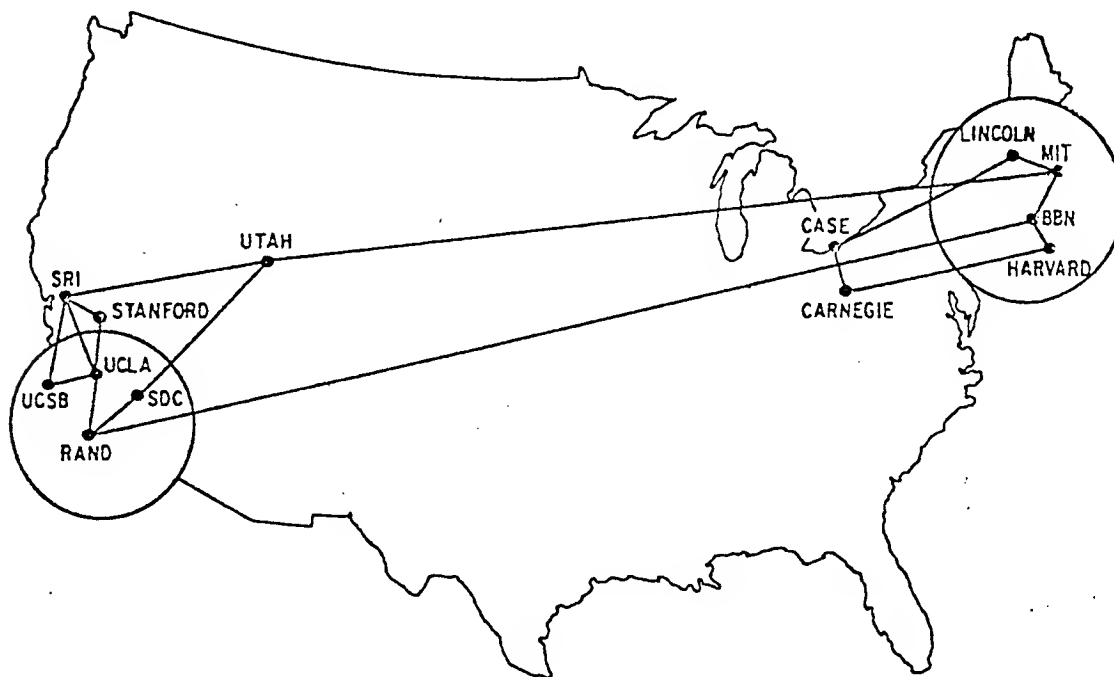


Figure 7: December 1970

By two-thirds of the way through 1971, two additional 516 IMPs had been installed, the prototype TIP was running at BBN, and two TIPs were operational within the network, at MITRE and AMES.

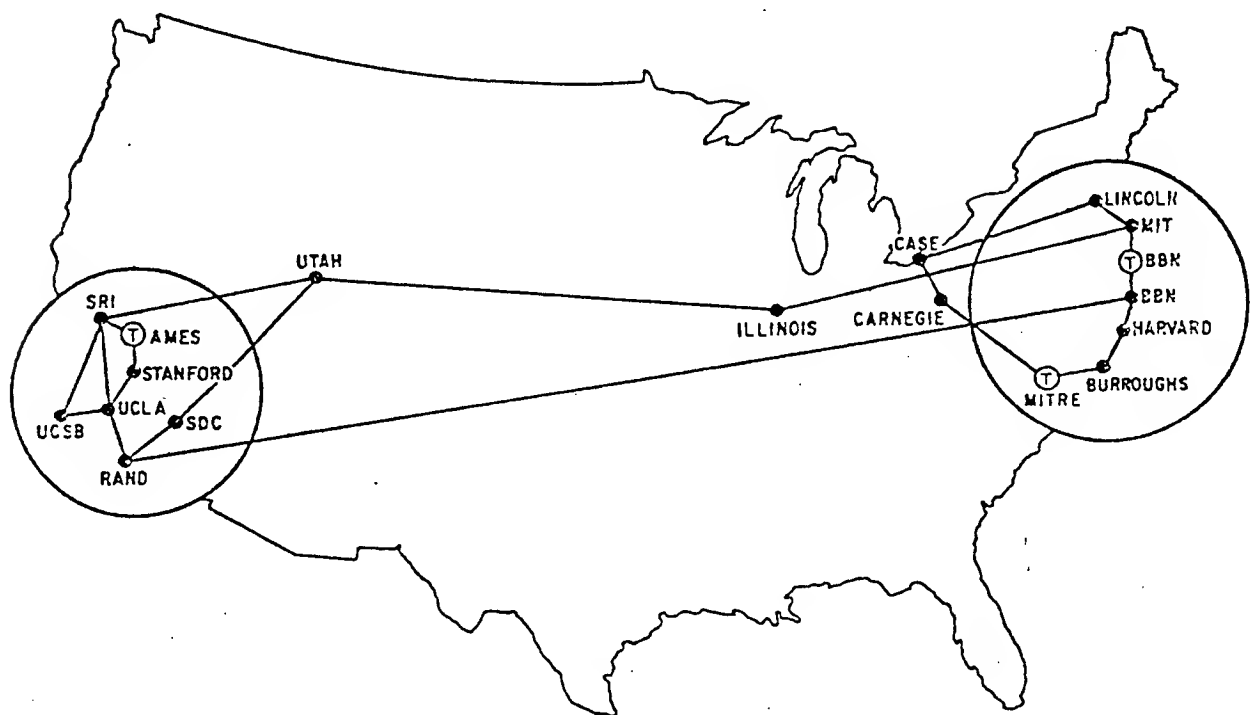


Figure 8: September 1971

The TIPs were based on Honeywell 316 instead of 516 computers and had as a component a 316-based IMP which was completely compatible with the 516 IMP but half as expensive. By early 1972 several additional IMPs and TIPs had been installed and the central part of the network between the East and West Coast clusters was beginning to fill out.

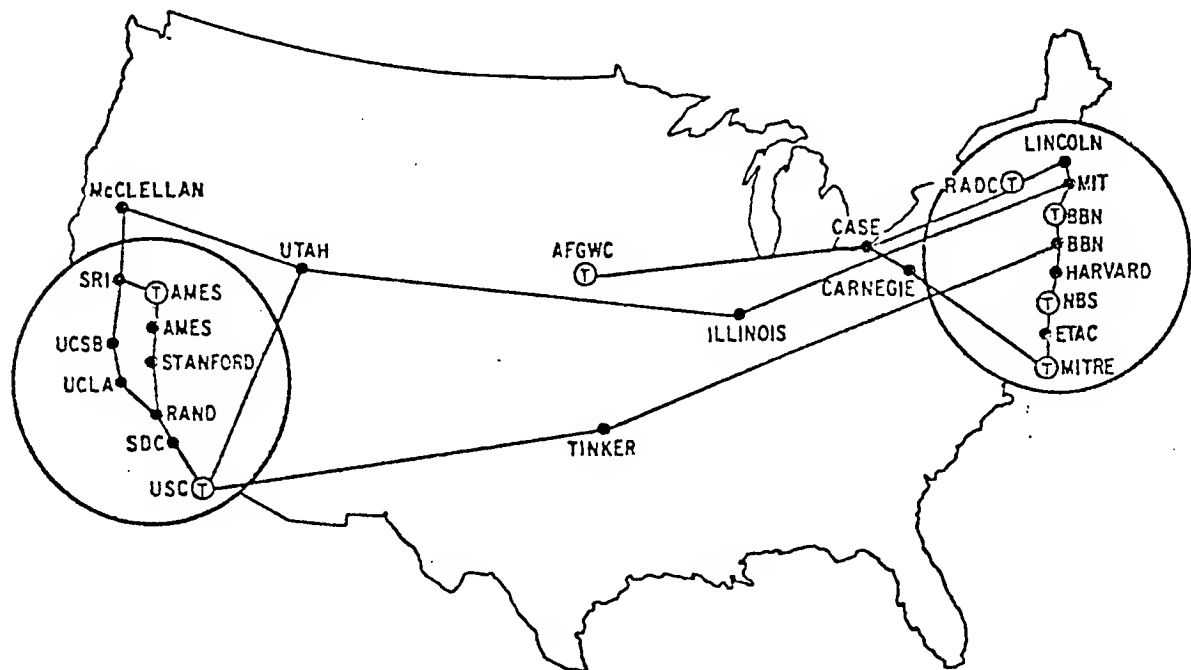


Figure 9: March 1972

By August 1972 a third cross-country line had been added and it was clear that in addition to the IMPs scattered throughout the center of the country, there were actually clusters of IMPs in four geographic areas, Boston, Washington. D.C., San Francisco, and Los Angeles.

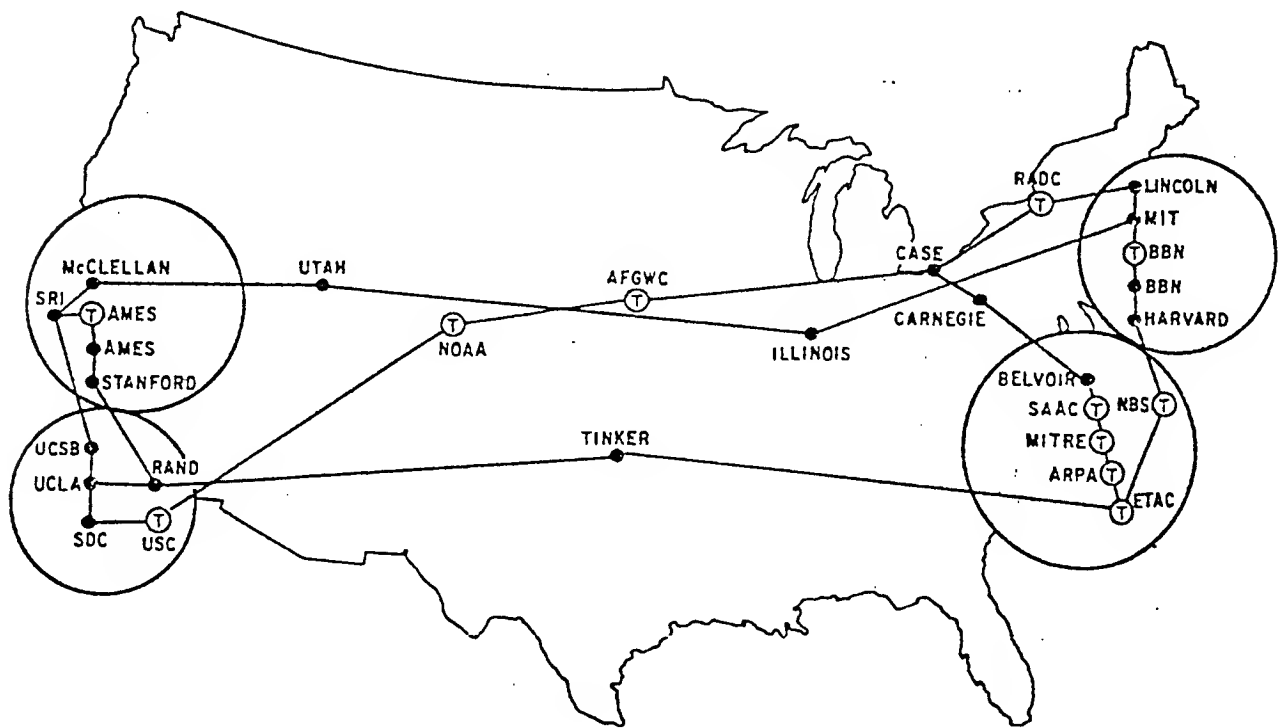


Figure 10: August 1972

There follow once-yearly maps for the years 1973 to 1977 with which the reader can follow the continued growth of the ARPANET topology.

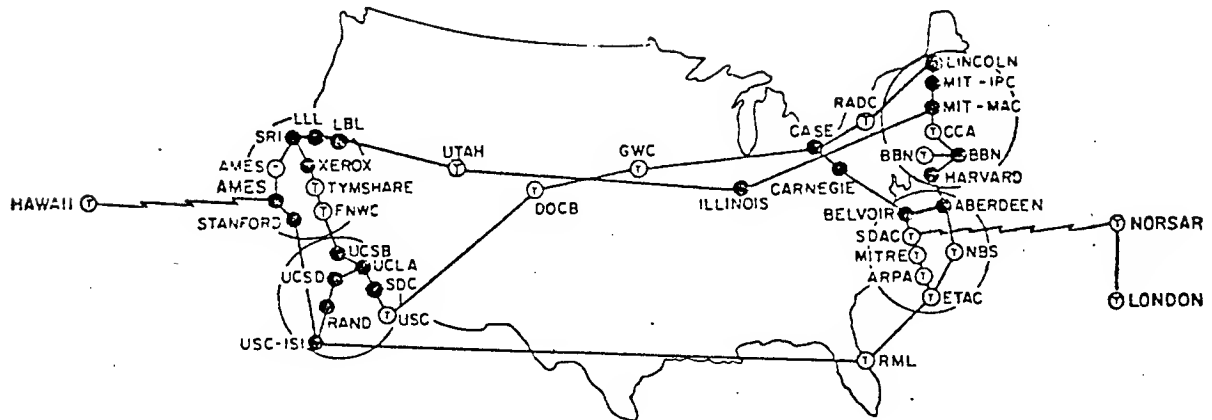


Figure 11: September 1973

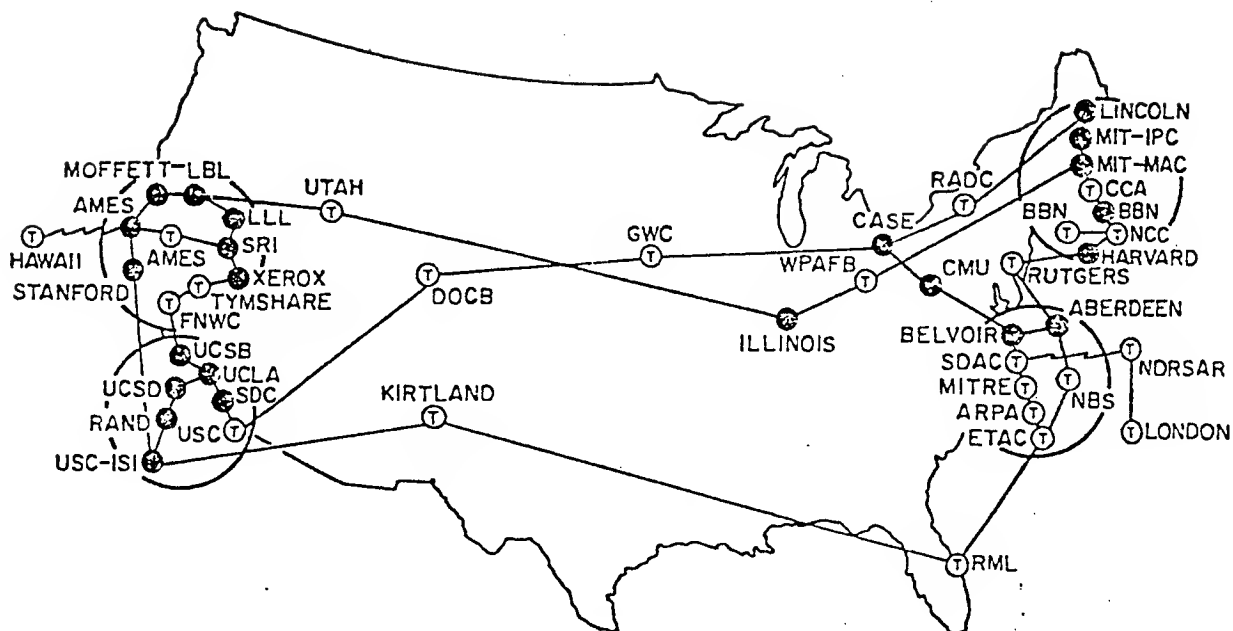


Figure 12: June 1974

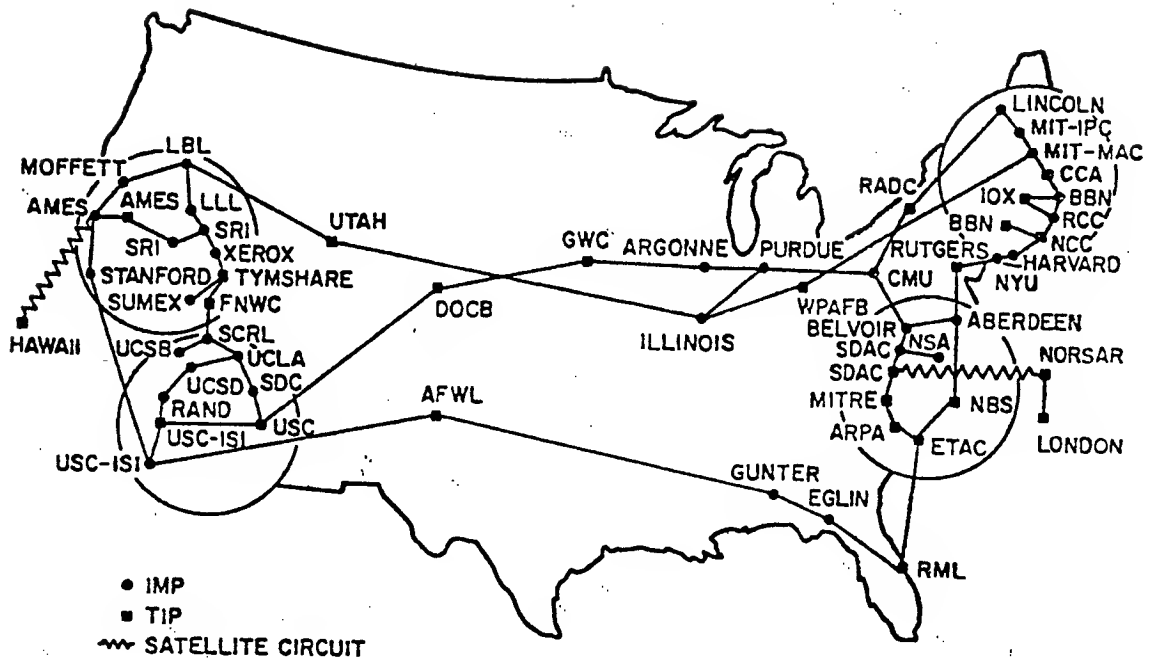


Figure 13: July 1975

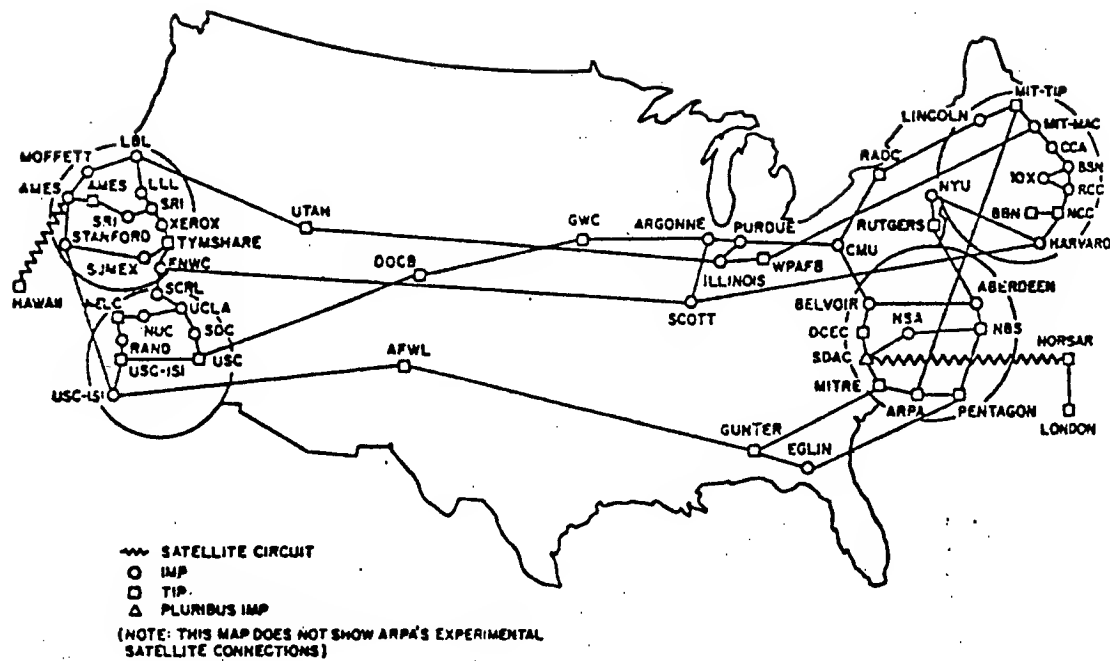


Figure 14: July 1976